An Empirical Investigation of Sampling and Other Errors in the National Stream Survey: Analysis of a Replicated Sample of Streams

W. Scott Overton
Stephen V. Stehman

Technical Report 119

October 1987

DEPARTMENT OF STATISTICS

Oregon State University

Corvallis, Oregon 97331

Copies available from: Department of Statistics, Oregon State University, Corvallis, OR 97331

DEPARTMENT OF STATISTICS
TECHNICAL REPORTS

2. Pierce, Donald A. and G. Nembauser. Computational results for a stopping rule problem on averages. February 1968.
4. Faulkenberry, G. David and James C. Daly. Sample sizes for tolerance limits on a normal distribution. August 1968.
5. Ramsey, Fred L. and Bradford R. Crain. Estimating spectral moments of a discrete, stationary time series. November 1968.
6. Pierce, Donald A. and J. L. Folks. Sensitivity of Bayes procedures to the prior distribution. September 1968.
7. Thomas, David R. Conditional locally most powerful rank tests for the two-sample problem with arbitrarily censored data. February 1969.
8. Bowman, V. Joseph and G. L. Nemhauser. Deep cuts in integer programming. February 1969.
9. Burnett, Thomas D. Estimation of stochastically varying regression parameters. March 1969.
10. Phillips, Charles J. and Donald Guthrie. Consistent empirical approximation of a priori distributions. June 1969.
11. Ramsey, Fred L. A small sample study of some non-parametric tests of location. July 1969.
13. Bowman, V. Joseph. Determining cuts in integer programming by two variable diophantine equations. August 1969.
14. Overton, W. Scott and Kenneth P. Burnham. A simulation study of live trapping and estimation of population size. August 1969.
15. Hartmann, Norbert A. F-max tables for mean squares with unequal degrees of freedom. September 1969.
16. Ramsey, Fred L., D. A. Pierce and V. J. Bowman. Triangularization of input-output matrices. October 1969.
21. Mullooly, John P. Small sample behavior of the stochastic first order reaction. April 1970.
23. Ramsey, Fred L. Bayesian Bioassay. May 1971.
24. Land, Charles E. and Bruce R. Johnson. A note on two-sided confidence intervals for linear functions of the normal mean and variance. July 1971.
25. Land, Charles E. Tables of confidence limits for linear functions of the normal mean and variance. February 1971.
26. Seely, Justus. Restricted maximum likelihood estimation for two variance components. March 1972.
27. Land, Charles E. Confidence interval estimation for functions of the normal mean and variance occurring in problems involving transformations of data. April 1972.
28. Pierce, Donald A. On some difficulties in a frequency theory of inference. April 1972.
29. Bogdanoff, David A. and Donald A. Pierce. Bayes-Fiducial inference for the Weibull distribution. April 1972.
30. Birkes, David, Yadolah Dodge, Norbert Hartmann, and Justus Seely. Estimability and analysis for additive two-way classification models with missing observations. July 1972.
32. Schroeder, Anita. Classification schemes on incomplete block models. August 1972.
33. Birkes, David, Y. Dodge, J. Seely. Estimability and analysis for additive three-way classification models with missing observations. September 1972.
34. Faulkenberry, G. David. Some results on prediction intervals. December 1972.
35. Brunk, H. D. and Donald A. Pierce. Note on Bayesian approaches to estimation of cumulative regression. January 1973.
37. Olsen, Anthony R. and Justus Seely. Invariant quadratic unbiased estimation for two variance components. June 1973.
38. Shepard, Lonnie B. and Donald Guthrie. Monte Carlo evaluation of empirical approximation of mixing distributions. June 1973.
39. Lembersky, Mark R. and Melvin L. Ott. A counterexample in continuous Markov decision chains. July 1973.
40. Brunk, H. D., D. R. Thomas, R. M. Elashoff, Calvin Zippin. Computer-aided prognosis. November 1973.
42. Brunk, H. D. and Donald A. Pierce. Large sample posterior normality of the population mean. June 1974.
43. Brunk, H. D. Uniform inequalities for conditional p-means given α-lattices. November 1974.

44. Seely, Justus. Minimal sufficient statistics and completeness for multivariate normal families. January 1975.

45. Birkes, David, H. D. Brunk, John W. Lee. Generalized exponential families. January 1975.

46. Pierce, Donald A. and Bradley R. Sands. Extra-Bernoulli variation in binary data. November 1975.

47. Scheurman, H. Lynn and K. Norman Johnson. Computational experience with a renewable resource (timber) harvest model. November 1975.

48. Lembersky, Mark R. Maximum return from a collection of standard and special appreciation assets. January 1976.

49. Brunk, H. D. and N. F. G. Martin. An inequality involving powers of sums of powers. January 1976.

50. Birkes, David and Justus Seely. Three-way classification designs of resolutions III, IV, and V. March 1976.

51. Brunk, H. D. Univariate density estimation by orthogonal series. March 1976.

52. Brunk, H. D. Stone spaces in probability. March 1976.

54. Jan, Shan Wu, David Birkes, and Roger Petersen. A test of hypothesis for a patterned correlation matrix: A Monte Carlo study. February 1976.

55. Pierce, Donald A. A random effects model for matched pairs of binomial data. September 1976.

56. Johansen, Søren. Homomorphisms and general exponential families. October 1976.

57. Jacroux, Michael and Justus Seely. Some sufficient conditions for establishing (M,S)-optimality. February 1977.

58. Thompson, Steve. The study of bear populations. June 1977.

59. Thomas, D. R. and D. A. Pierce. Neyman's smooth goodness-of-fit test when the hypothesis is composite. April 1977.

60. Ramsey, Fred L. and J. Michael Scott. Use of circular plot surveys in estimating the density of a population with Poisson scattering. January 1978.

61. Kopecky, Kenneth J. and Donald A. Pierce. Efficiency of smooth goodness-of-fit tests. April 1978.

62. Brunk, H. D. Models for integration. July 1978.

63. Lindstrom, F. Tom. Function minimization by the Nelder-Mead simplex method. October 1978.

64. Lindstrom, F. T. and D. S. Birkes. On the estimation of parameters in a non-linear one compartment stochastic open model by apprximate likelihood functions. January 1979.

65. Ramsey, Fred L. Parametric models for line-transect surveys. February 1979.

66. Dodge, Y. and D. R. Thomas. On the performance of non-paramtric and normal theory multiple comparison procedures. February 1979.

67. Pierce, Donald A., William A. Stewart and Kenneth J. Kopecky. Distribution-free regression analysis of grouped survival data. August 1979.

68. Brunk, H. D. Bayesian least squares estimates of univariate regression functions. August 1979.

69. Lindstrom, F. T. An objective function surface exploration algorithm. August 1979.

70. Arthur, J. A comparison of optimization methods on a non-linear regression problem. August 1979.

71. Brunk, H. D. Estimation of stimulus-response curves by Bayesian least squares. November 1979.

72. Seely, Justus. Parametrizations and correspondence in linear models. November 1979.

73. Lindstrom, F. Tom. FUNMIN: An online-interactive mixed simplex-Marquardt least squares or likelihood minimization algorithm. January 1980.

74. Kafadar, Karen. An empirical investigation of small samples from symmetric populations for constructing robust confidence intervals. May 1980.

75. Chou, Chung-Kuang and David A. Butler. Hazardous-inspection initiation policies. June 1980.

76. Kafadar, Karen. Borrowing scale information in estimating location robustly. June 1980.

77. Butler, David A. and Gerald J. Lieberman. An early-accept modifcation to the test plans of military standard 781C. June 1980.

78. Rosenfield, Donald B., Roy D. Shapiro and David A. Butler. Optimal strategies for selling an asset. December 1980.

79. Dodge, Y. and F. T. Lindstrom. An alternative approach to least squares estimation when dealing with contaminated data. April 1981.

80. Brunk, H. D. Boolean universes for stochastic models. December 1981.

81. Butler, David A. and G. J. Lieberman. Inspection policies for fault location. September 1982.

82. Thompson, Steven K. and Fred L. Ramsey. Adaptive sampling of anaimal populations. March 1983.

83. Hussein, Abou-Bakr A. and David R. Thomas. Optimal design of np-charts. March 1983.

84. Shih, John and H. D. Brunk. Bayes least squares linear estimation of densities. March 1983.

85. Ramsey, Fred L. and Christopher P. Marsh. Diet dissimilarity. May 1983.

86. Seely, Justus and David Birkes. Parametrizations and resolution IV. May 1983.

87. Schafer, Daniel W. The adequacy of least squares before and after correction for attenuation. June 1983.

88. Seely, Justus. Some continuity properties of dispersion matrices of estimators in linear models. August 1983.

89. Brunk, H. D. Bayes least squares linear regression is asymptotically full Bayes: Estimation of special densities. November 1983.

90. Lee, Youngjo. The unbiased estimation of risk for discontinuous estimation of the mean of the multivariate normal distribution. November 1983.

91. Lee, Youngjo and O. Kempthorne. An estimation procedure for K-nested linear models. November 1983.

92. Lee, Youngjo. Asymptotic properties of the shrinkage estimation. November 1983.

93. Stark, Steven B. A model study of the alfalfa leafcutter bee - seed production system. November 1983.

94. Schafer, Daniel W. Combining information for measurement error model regression. January 1984.

95. Gray, Robert J. and Donald A. Pierce. Goodness-of-fit tests for censored survival data. January 1984.

96. Pierce, Donald A. and Dale L. Preston. Hazard function modelling for dose-response analysis of cancer incidence in the A-bomb survivor data. March 1984.

97. Pierce, Donald A., Dale L. Preston and Toranosuke Ishimaru. A method for analysis of cancer incidence in Japanese A-bomb survivors, with application to acute leukemia. March 1984.

98. Lee, Youngjo. A short note on the positive-part shrinkage estimator for the location parameter under arbitrary convex loss function. March 1984.

99. Rossi, Richard J. and F. L. Ramsey. P-values for extreme values of the normal, student's t, chi square and F statistics. May 1984.

100. Arthur, Jeffrey L. Vector maximization by efficient directions. August 1984.

101. Thomas, Dana L. and David R. Thomas. Confidence bands for percentiles in the linear regression model. August 1984.

102. Brunk, H. D. Coherence and consistency of inferences. September 1984.

103. Carter, Nancy J. and G. David Faulkenberry. Predicting unit variate values in a finite population. December 1984.

104. Limam, Mohamed and David R. Thomas. Simultaneous tolerance intervals for the linear regression model. February 1985.

105. Birkes, David R., Cliff B. Pereira and Justus F. Seely. Admissibility among affine sets of linear estimators under quadratic loss. March 1985.

106. Wildman, Valerie Jean and Fred L. Ramsey. Estimating effective area surveyed with the cumulative distribution function. May 1985.

107. Limam, Mohamed and David R. Thomas. Simultaneous tolerance intervals in the random one-way model with covariates. August 1985.

108. Pierce, Donald A. and Dale L. Preston. Analysis of cancer mortality in the atomic bomb survivor cohort. September 1985.

109. Brunk, H. D. The compact ramification of a set, and induced measures. November 1985.

110. Schafer, D. W. Covariate measurement error in generalized linear models. November 1985.

111. Birkes, David and Yadolah Dodge. The number of minimally connected block designs. January 1986.

112. Brunk, H. D. Fully Coherent Inference. June 1986.

113. Overton, W. Scott. Working Draft, Analysis Plan for the Eastern Lake Survey, March 8, 1985. July, 1986.

114. Overton, W. Scott. A sampling plan for streams, in the national surface water survey. July 1986.

115. Overton, W. Scott. Phase II Analysis Plan, National Lake Survey -- Working Draft, April 15, 1987. April 1987.

116. Arthur, Jeffrey L., James O. Frendewey and Robert T. Sumichrast. GENLP: A FORTRAN Routine for Generating Linear Programs with Known Optimal Solutions. May 1987

117. Overton, W. Scott. Sampling and Analysis Plan for Streams in the National Surface Water Survey. June 1987.

118. Stehman, Stephen V. and W. Scott Overton. An empirical investigation of the variance estimation methodology prescribed for the National Stream Survey: Simulated sampling from stream data sets. October 1987.

119. Overton, W. Scott and Stephen V. Stehman. An empirical investigation of sampling and other errors in the National Stream Survey: Analysis of a replicated samples of streams. October 1987.

120. Rossi, R. and H. D. Brunk. L_1 and L_2 cross-validation for density estimation with special reference to orthogonal expansions. November 1987.

121. Brunk, H. D. and P. W. Jones. Fitting conditional distributions using simulated investigators. November 1987.

122. Butler, David A. Bitwise Simulation of the Reliability of a Coherent System. February 1988.

An Empirical Investigation of Sampling and Other Errors in the National Stream Survey; Analysis of a Replicated Sample of Streams

by W. Scott Overton
and Stephen V. Stehman

Department of Statistics, Oregon State University
and Biometrics Unit, Cornell University

Technical Report 119
Department of Statistics, Oregon State University

October 1987

This paper is a contribution of the Aquatic Effects Research Program, funded by the U. S. Environmental Protection Agency, through the National Acid Precipitation Assessment Program. This paper has not been subjected to EPA's peer and policy review, and therefore does not necessarily reflect the views of the Agency.

INTRODUCTION

Design of the Environmental Protection Agency's National Stream Survey (Messer et al, 1986; Overton, 1985; Overton, 1987), involved several components that required validation in establishing the validity of the survey. Elsewhere we have reported theoretical development (Overton, 1987), theoretical and simulation studies (Stehman and Overton, 1987a), and simulation studies (Stehman and Overton, 1987b) that establish validity of specific aspects of this design. The present paper presents the last of these investigations, being a study of the behavior of the method under operational selection of samples from the map materials used in the survey.

Specifically, design elements that require verification are:

1) Choice of variance estimator; the Horvitz-Thompson (1952) variance estimator is known to exhibit undesirable properties in some circumstances. There are several theoretical reasons for which this v—estimator seems best for the National Stream Survey (NSS), but empirical evidence of its adequacy was desired.
2) Use of any variance estimator for this problem requires approximation to the 2nd order inclusion probabilities. This issue has been discussed in depth elsewhere (Stehman and Overton, 1987b), and it is sufficient here to say that the approximation used has been shown to be better than the available alternatives in the circumstance of the NSS.
3) The first order inclusion probability is conceptually well identified, but is subject to several sources of error in measurement. Specifically, $\pi_i = a_{1i}/Q$, where Q is the area on the frame maps corresponding to the spacing on the dot-grid overlay, and a_{1i} is the direct watershed area of the i^{th} reach. Q is taken at its nominal value of 64, and a_{1i} is measured on maps by planimeter. Errors in either of these components can generate bias.
4) Somewhat subjective considerations are involved in identifying the reach associated with a grid-dot. It is conceivable that these could lead to problems.

The other studies have addressed the first two of these issues, but are fundamentally incapable of assessing the latter two. The study reported here will serve to verify the results of the other studies, and in addition shed light on the issues of selection bias and measurement errors in the π's. More generally, the current study addresses the general issues of operational application of the design and behavior of the specified estimators.

THE REPLICATED STUDY

Ten independent replicated samples were generated by the operational sampling protocol from the Knoxville (Tenn) quadrangle; this area was part of the pilot study area of the NSS (Messer, et al, 1986). For comparison, the entire population of target reaches in this study area was identified, and key variables determined for each. This population tally provides the basis for assessment of accuracy of estimates generated from the ten samples, and also of accuracy of estimated precision.

Although 10 replicates are inadequate to establish performance characteristics, they are sufficient for verification of behavior that has been inferred on other grounds. Elsewhere, we have used simulation experiments extensively in this verification (Stehman and Overton 1987a and b) with computer selection of samples from populations composed of reach data from the NSS. The replicated samples analyzed here have been selected by human samplers using the sampling protocol of the National NSS, and have the role of confirming that the behavior of the sampling strategy, as realized by that protocol of application, is essentially as expected. Problems of application can show up here, but not be seen in the simulations performed on the computer. Additional verification is provided by computer simulation on the fully tallied population, allowing a larger sample than was possible in the more intensive simulation studies.

The replications required one difference in protocol from that used in the operational NSS; in order that replications could be considered independent, placement of the grid was randomly selected, and carefully identified. Otherwise, all selection was made by the established survey protocol. Each of five persons selected two samples; reps 1 and 2, reps 3 and 4, . . . , reps 9 and 10 were selected by the five samplers.

The selection protocol is detailed by Messer et al (1986), and by Allen et al (1987). In brief, a regular 64-sq.-mi.-per-dot square, dot-grid randomly overlays the map. Associated with each dot there is identified either a reach or no-reach, and a set of site rules determine whether a selected reach is in the identified target population. Essentially, a reach is selected if a grid-dot falls in the direct watershed of that reach.

Replicates were verified against the total population tally, and all discrepancies resolved by reexamination of the maps. This cross-sample validation also represents a step not present in the operational survey, and the errors so eliminated may reflect errors that would not be detected in the operational survey. The purpose of this refinement was to concentrate in this study on errors of selection and interpretation, to the exclusion of gross errors of measurement or application of the site rules, which may be controlled by rigorous QA/QC[1] protocol. The human element of sampling is a central focus of this study. Note, however, that measurement errors in the a_i's are still present, and constitute an important dimension of the concern.

Table 1a Physical properties of the 10 replicates.

Rep	Points in Study Area	Target reaches Total	First Order	Second Order	Higher Order
1	126	98	75	6	17
2	135	95	73	8	14
3	135	100	79	7	14
4	135	95	71	8	16
5	126	101	80	12	9
6	120	92	69	12	11
7	135	102	82	11	9
8	126	91	75	9	7
9	126	99	74	14	11
10	126	95	74	10	11
Mean		96.8	75.2	9.7	11.9
SD		3.77	4.05	2.54	3.25

In Table 1a, it is seen that the pattern of variation in total number of points in the study area is somewhat unusual; only three values are represented in the ten samples, and one of those only once. This phenomenon is due to coincidence of the regular pattern of the dot-grid and the regular outline of the study area. The outline of designated study areas in the operational survey are not so regular, and such patterns would be much less pronounced. Note that the numbers of target reaches in the replicated study does not show such pattern.

[1]Quality Assurance and Quality Control protocol for data collection and data management.

On these samples, each reach is characterized by three attributes, r, a_i and l, being reach order, area of direct watershed (sq.mi.) and reach length (mi.), respectively. The complete population tally identified these attributes for each of the 1296 reaches, Table 1b.

Table 1b. Characteristics of the full Target Reach Population.

	Total by Order				Pop. Variance[a]	
	All	First	Second	Higher	All	First
Number of reaches	1296	909	169	218	0.0	0.2096
Area of direct watershed	5940.8	4542.6	616.2	782.0	17.28	17.42
Reach length	4192.1	3184.2	443.8	564.1	4.354	5.429

[a] The standard finite population parameter, $V_y = \frac{1}{N-1}\sum(y-T_y/N)^2$.

ESTIMATES OF POPULATION ATTRIBUTES

Population attributes are estimated from each of the ten replicates, as summarized in Tables 2a, 2b, and 2c. The Horvitz-Thompson estimates are used throughout. Standard errors are square roots of variances estimated by the Horvitz-Thompson variance formula. All of these are standard specifications for the Surface Water Surveys. Details and derivation are provided by Overton, 1985, 1987.

$$\hat{T}_y = \sum y/\pi = \sum wy = 64\sum y/a_i \, , \tag{1}$$

where a_i is the area of the direct watershed of the unit reach,

$w = 64/a_i = 1/\pi$,

y is any reach attribute of interest,

and π is the inclusion probability of that unit reach.

$$\hat{V}(\hat{T}_y) = \sum_i y^2 w(w-1) + \sum_i\sum_{j\neq i} y_i y_j (w_i w_j - w_{ij}) \, . \tag{2}$$

where w_{ij} is the inverse joint inclusion probability of units i and j. Summations are over subsets of the sample belonging to the population under consideration.

The form of special cases of the estimators shed light on the results:

a) Let $y = 1$, so that T_y is the total number of target reaches, N;

then $\hat{N} = 64\sum 1/a_i$. (3)

b) Let $y = a_1$, so that T_y is the total area of direct waterashed of target reaches, A;

then $\hat{A} = 64n$, (4)

where n is the number of target reaches in the sample.

c) Let $y = l$, so that T_y is the total length of target reaches, L;

then $\hat{L} = 64\sum l/a_1$. (5)

These quantities are calculated for each replication and variable, and summarized in subsequent tables. Standard errors (SE) are root-variances. Statistics are also calculated among the reps, as for example, standard deviations (SD). The standard deviation of the replicated estimates is an estimate of the parameter being estimated by the standard errors, and the SD is the standard against which the mean SE is to be judged. The mean of the rep estimates contrasts to the true value of population totals from Table 1b.

Table 2a. Estimates and Standard Errors for the 10 reps.

	Number of Target Reaches					
	All Reaches		First Order		Higher Order	
Rep	Est	SE	Est	SE	Est	SE
1	1,302	116.3	924	98.2	378	95.2
2	1,787	273.7	1,015	112.2	771	270.0
3	2,188	724.6	1,185	183.5	1,003	713.0
4	2,350	798.8	954	119.3	1396	801.8
5	1,413	127.6	1,168	128.4	245	63.7
6	1,455	174.2	957	128.8	497	145.9
7	1,517	157.9	1,028	103.9	489	145.6
8	1,354	151.8	1,134	149.9	220	65.4
9	1,272	109.7	1,041	115.1	231	47.9
10	1,195	113.4	915	103.3	279	77.3
Mean	1,583	274.8	1,032	124.3	551	242.6
SD	397.8	261.6	99.9	25.8	392.0	279.6
True	1,296		909		387	
eliminating reps 3 and 4 . . .						
Mean	1,412	153.1			389	113.9
SD	183.5	54.1			190.5	73.0

There is an apparent tendency to overestimate each of the three parameters; the only quantity underestimated is watershed area of higher order reaches. Variance is well estimated for reach length and for number of first order reaches, overestimated for watershed area, and underestimated for number of higher order reaches, and hence for total number of reaches. In Table 2a, reps 3 and 4 stand out as quite different from the rest. To emphasize this difference, the mean and standard deviation of the other 8 reps are also given. It is clear that these two reps are responsible for a considerable part of the bias and a considerable part of the variance of this set of estimates.

Table 2b. Estimates and Standard errors for the 10 reps.
Direct Watershed Area (sq mi)

	All Reaches		First Order		Higher Order	
Rep	Est	SE	Est	SE	Est	SE
1	6,272	281.4	4,800	331.3	1,472	264.0
2	6,080	321.0	4,672	349.3	1,408	262.1
3	6,400	307.7	5,056	345.1	1,344	256.7
4	6,080	319.1	4,544	347.1	1,536	271.1
5	6,464	268.8	5,120	325.2	1,344	249.2
6	5,888	277.8	4,416	322.5	1,472	263.8
7	6,528	300.9	5,248	340.1	1,280	254.1
8	5,824	302.2	4,800	330.7	1,024	226.1
9	6,336	276.8	4,736	333.5	1,600	265.7
10	6,080	288.8	4,736	329.4	1,344	251.6
Mean	6,195	294.5	4,813	335.3	1,382	256.4
SD	241	18.4	259.2	9.4	160.2	12.7
True	5,941		4,543		1,398	

Table 2c. Estimates and Standard errors for the 10 reps.
Reach Length (miles)

	All Reaches		First Order		Higher Order	
Rep	Est	SE	Est	SE	Est	SE
1	4,437	273.9	3,377	287.4	1,061	210.1
2	4,706	341.5	3,331	291.1	1,375	307.9
3	5,184	406.6	3,912	342.2	1,272	340.8
4	4,439	358.8	3,063	278.2	1,375	329.3
5	4,551	264.1	3,620	283.9	930	192.3
6	4,534	317.7	3,372	326.3	1,162	233.0
7	5,002	331.2	3,903	318.7	1,098	258.3
8	4,356	344.9	3,576	343.2	780	203.8
9	4,407	266.9	3,481	297.0	926	169.7
10	4,155	279.6	3,182	280.6	973	208.7
Mean	4,577	318.7	3,481	304.8	1,095	245.3
SD	309.6	47.1	279.2	25.3	201.0	60.8
True	4,196		3,184		1,008	

The above effect is not seen in Table 2b, in estimation of total direct watershed area. In fact, this estimate is unaffected by the value of a_1, and is simply a function of the number of target reaches in the sample, as is seen from Eqn.4. Estimation of total reach length, Table 2c, follows closely to this pattern, with only slight inflation apparent in reps 3 and 4. This is probably due to the high correlation (ρ = .83) between a_1 and l. It is also notable that each of these attributes shows slightly biased estimates, and it seems that something in addition to the small a_1's is contributing to bias.

It is of interest, then, to inspect reps 3 and 4 for content that is causal for this overestimate of numbers and the variance of estimated numbers. A clue to the cause is provided by Table 3.

Table 3a. Frequency of target reaches, by size of direct watershed (square miles).

Size	Replication										
Interval	1	2	3	4	5	6	7	8	9	10	Total
(0.0,0.2]	0	0	1	2	0	0	0	0	0	0	3
(0.2,0.5]	0	2	1	0	0	0	0	0	0	0	3
(0.5,1.0]	0	2	0	0	0	2	2	2	0	0	8
(1.0,2.0]	4	7	4	9	10	7	4	8	8	4	65
(2.0,3.0]	14	12	14	6	10	10	15	8	7	16	112
(3.0,4.0]	7	7	9	11	10	8	12	7	14	6	91
(4.0,5.0]	10	13	14	9	11	11	9	11	6	5	99
(5.0, 10.0]	38	32	34	31	34	25	31	31	35	36	327
(10.0,]	25	20	23	27	26	29	29	24	29	28	260

Table 3b. Frequency of headwater reaches by size of direct watershed (square miles).

Size	Replication										
Interval	1	2	3	4	5	6	7	8	9	10	Total
(0.0,0.2]	0	0	0	0	0	0	0	0	0	0	0
(0.2,0.5]	0	0	1	0	0	0	0	0	0	0	1
(0.5,1.0]	0	0	0	0	0	1	0	2	0	0	3
(1.0,2.0]	2	5	2	8	8	5	1	7	8	2	48
(2.0,3.0]	9	10	12	3	10	6	13	6	6	14	89
(3.0,4.0]	6	7	7	8	7	6	9	5	13	5	73
(4.0,5.0]	10	10	12	6	9	7	8	9	5	4	80
(5.0, 10.0]	27	25	26	25	29	21	25	27	23	28	256
(10.0,]	21	16	19	21	17	23	26	19	19	21	202

In Table 3a, reps 3 and 4 are unique in having reaches in the smallest class, less than .2 sq mi. Rep 3 also has a reach in the next smallest class, as does Rep 2; inspection of Table 2a reveals that rep 2 also exhibits somewhat large estimate and SE. It is the nature of these estimators that the weight associated with a sample unit is large if the a_i is small, and this is the source of the identified characteristic of reps 3 and 4 (and also 2). However, some deliberation is indicated in identifying the nature of these properties, and the appropriate treatment.

If very rare selections are made in the appropriate frequency, and if the a_i's are accurately measured, then no bias is associated with occurrence of such rare units. However, a sample that happens to contain a rare element may have a quite large error, and so be improved, in the sense of smaller MSE, by an appropriate modification of the estimator. This modification would opt for a small increase in bias in return for a large decrease in variance. Such a modification is provided by the simple device of "scoring" a too-small value of a_i to an acceptable value, as we will explore in Tables 4a, 4b, 4c, and 4d.

But if supposedly rare selections occur with too-great frequency, then bias can result, in addition to inflated variance, and in such a circumstance, the case for scoring is even stronger. If variance inflation by too-small a_i's is an issue, then it may be possible at the design stage to redefine the sampling units in a manner so as to eliminate the offending units; such a redefinition would clearly have been feasible for the NSS. But the analytic device of scoring is effective, as will be demonstrated, so the design issue is not crucial.

Table 4a. The effect of scoring on estimated number of reaches — *target reaches.*

	No Scoring		Score to 0.2		Score to 0.5		Score to 1.0	
Rep	Est	SE	Est	SE	Est	SE	Est	SE
1	1,302	116.3						
2	1,787	273.7			1,704	226.6	1,547	160.9
3	2,188	724.6	1,797	361.3	1,577	203.5	1,449	139.9
4	2,350	798.8	1,902	454.5	1,518	207.7	1,390	145.7
5	1,413	127.6						
6	1,455	174.2					1,401	149.4
7	1,517	157.9					1,474	139.4
8	1,354	151.8					1,330	141.5
9	1,272	109.7						
10	1,195	113.4						
Mean	1,583	274.8	1,499	204.0	1,431	158.9	1,377	134.4
SD	397.8	261.6	246.6	119.6	154.6	42.8	104.1	16.9
True	1,296							

Scoring will be explored at three levels, .2, .5 and 1.0 sq mi. At each level, all values of a_i below the designated value will be set to that value. By inspection of Table 3a, it is seen that of the 968 sample reaches in the 10 reps, only 3 are affected by scoring to .2, 6 by scoring to .5 and 14 by scoring to 1.0. The results are given in tables 4a, 4b, 4c, and 4d. Recall that estimated direct watershed area is not affected by scoring, so only estimated numbers of reaches and reach length are analysed.

Table 4b. The effect of scoring on estimated number of reaches — *headwater reaches, only.*

	No Scoring		Score to 0.2		Score to 0.5		Score to 1.0	
Rep	Est	SE	Est	SE	Est	SE	Est	SE
1	924	98.2						
2	1,015	112.2						
3	1,185	183.5			1,157	161.7	1,093	121.9
4	954	119.3						
5	1,168	128.4						
6	957	128.8					952	126.2
7	1,028	103.9						
8	1,134	149.9					1,110	139.1
9	1,041	115.1						
10	915	103.3						
Mean	1,032	124.3	1,032	124.3	1,029	122.1	1,020	116.6
SD	99.9	25.8	99.9	25.8	95.4	20.6	85.0	12.8
True	909							

Table 4c. The effect of scoring on estimated reach length — *target reaches.*

	No Scoring		Score to 0.2		Score to 0.5		Score to 1.0	
Rep	Est	SE	Est	SE	Est	SE	Est	SE
1	4,437	273.9						
2	4,706	341.5			4,642	325.1	4,499	300.7
3	5,184	406.6	5,040	346.8	4,944	330.0	4,863	321.1
4	4,439	358.8	4,307	310.4	4,209	300.8	4,176	301.2
5	4,551	264.1						
6	4,534	317.7					4,483	309.5
7	5,002	331.2					4,939	316.0
8	4,356	344.9					4,304	324.6
9	4,407	266.9						
10	4,155	279.6						
Mean	4,577	318.7	4,549	307.7	4,524	303.4	4,481	295.8
SD	309.6	47.1	289.6	33.7	279.8	30.2	258.0	22.8
True	4,192							

The replicates influenced by scoring have been indicated by Table 3, and the magnitude of the change and the effect on mean estimate, on estimated variance, and on the variance among reps, are seen in Table 4 for the several parameters. Briefly, scoring has pronounced effect on estimation of N, the total number of reaches, but very little effect on estimation of numbers of headwater reaches. This is because few headwater reaches have small a_1's, and the implication is that great effect is seen on estimation of numbers of reaches of order higher than 1, these having most of the small a_1's. Scoring has little effect on estimation of reach length; the same replicates are changed, but the magnitude of the change is minor.

Table 4d. The effect of scoring on estimated reach length — *headwater reaches, only.*

	No Scoring		Score to 0.2		Score to 0.5		Score to 1.0	
Rep	Est	SE	Est	SE	Est	SE	Est	SE
1	3,377	287.4						
2	3,331	291.1						
3	3,912	342.2			3,887	334.8	3,830	324.8
4	3,063	278.2						
5	3,620	283.9						
6	3,372	326.3					3,362	323.2
7	3,903	318.7						
8	3,576	343.2					3,524	321.8
9	3,481	297.0						
10	3,182	280.6						
Mean	3,481	304.8	3,481	304.8	3,479	304.1	3,467	300.7
SD	279.2	25.3	279.2	25.3	275.0	24.3	264.7	19.2
True	3,184							

Although scoring has been conceived as a method of variance reduction, at the cost of increased bias, it is seen from Table 4 that bias reduction has resulted in addition to variance reduction. This implies that some form of selection bias is being incurred in the operational selection of samples; rare events are apparently occuring with too-great frequency. This sort of bias can be identified only from replicates of the kind analysed here, and further investigation into the possible nature of such selection is indicated.

The analyses to explore this are similar to goodness of fit analyses, but modified to account for the variable probability of selection, and focussed on the kinds of estimates used in the survey, rather than on the more usual statistic. In Table 5a, the variable, n, is identified as the number of sample units, totaled over the ten reps, that are in the several size classes determined by a_1. The usual analysis would compare this statistic to its expected value, given the population structure and the theoretical sampling model. Instead, the same information is provided by identifying the total population estimate, $\hat{A}$, from this number of sample reaches, to be compared to the total direct watershed area in the various classes. Note that $\hat{A}$ is obtained from n by multiplying by 64/10. The comparison is simply scaled in familiar units.

Also in Table 5a, the statistic relating to the estimated number of reaches is identified, $\Sigma(1/a_{1i})$, and this could be compared to the expected value of that statistic. Instead, the estimated numbers ($\hat{N}$) are compared against the established population parameter. Again, $\hat{N}$ is obtained from the statistic by multiplying by 6.4. For numbers of reaches, the estimate based on a_1's scored to .5 is given in addition to the unscored estimate. Table 5b repeats these analyses for headwater reaches.

Table 5a. Comparison of the basic statistics, and associated estimates, by size class of reach, with the population value of the parameter — *all target reaches.*

	DIRECT WATERSHED AREA				NUMBER OF REACHES				
CLASS	n	$\hat{A}$	$\sum_P a_1 - A$	Δ	$\sum_S 1/a_1$	$\hat{N}$	$\hat{N}_s$	N	Δ
0—.2	3	19.2	2.72	16.5	28.10	179.84	38.40	22	16.4
.2—.5	3	19.2	13.76	5.4	7.74	49.54	38.40	38	.4
.5—1	8	51.2	52.67	-1.5	10.35	66.24	66.24	69	-2.8
1—2	65	416.0	305.51	110.5	41.36	264.70	264.70	199	65.7
2—3	112	716.8	625.81	91.0	45.72	292.61	292.61	253	39.6
3—4	91	582.4	555.46	26.9	26.27	168.13	168.13	160	8.1
4—5	99	633.6	624.19	9.4	22.24	142.34	142.34	140	2.3
5—10	327	2092.8	2068.60	24.2	47.66	305.02	305.02	301	4.0
10<	260	1664.0	1692.09	-28.1	17.95	114.88	114.88	114	0.9
TOTAL	968	6195.2	5940.81	254.4	247.39	1583.30	1430.72	1296	134.7

Table 5b. Comparison of the basic statistics, and associated estimates, by size class of reach, with the population value of the parameter — *headwater reaches.*

	DIRECT WATERSHED AREA				NUMBER OF REACHES				
CLASS	n	$\hat{A}$	$\sum_P a_1 - A$	Δ	$\sum_S 1/a_1$	$\hat{N}$	$\hat{N}_s$	N	Δ
0—.2	0	0.0	.23	-.2	0.00	0.00	0.00	2	-2.0
.2—.5	1	6.4	1.03	5.4	2.44	15.62	12.80	3	9.8
.5—1	3	19.2	13.68	5.5	3.46	22.14	22.14	17	5.1
1—2	48	307.2	211.67	95.5	29.38	188.03	188.03	133	55.0
2—3	89	569.6	484.88	84.7	36.11	231.10	231.10	195	36.1
3—4	73	467.2	451.07	16.1	21.20	135.68	135.68	130	5.7
4—5	80	512.0	499.14	12.9	18.03	115.39	115.39	112	3.4
5—10	256	1638.4	1603.71	34.7	37.31	238.78	238.78	233	5.8
10<	202	1292.8	1277.22	15.6	13.38	85.63	85.63	84	1.6
TOTAL	752	4812.8	4542.63	270.2	161.31	1032.38	1029.56	909	120.6

Inspection of estimated watershed area, $\hat{A}$, and numbers of target reaches, $\hat{N}$, by size class of direct watershed, reveals several interesting patterns. The smallest class evidences the greatest error in estimated N (179.84 — 22), with the estimate reduced to 38.4 by scoring. The classes 1—2 and 2—3 also have substantial error in estimates of both parameters. Errors are really quite small for the four largest classes, and for the second and third. Overall, there is a general tendency to overestimate both A and N, and some effort is justified in the attempt to understand this error and possibly to modify the estimators to account for it. Scoring has appreciable effect only on class 0—.2.

The observed bias in the smallest class is clearly due to some sort of irregularity in selection or identification of very small watersheds. Elsewhere this has been examined with respect to effect on estimated variance; variance and the estimate of variance increase when these very small watersheds are included in the sample. But here it is indicated that these small watersheds are included at six times their expected frequency. But the actual numbers involved are small, with perhaps 3 excess small a_1 reaches in the 10 reps. This would clearly not be significant if they were not apparently confined to the samples (3 and 4) collected by a single sampler. This increases the likelihood that the effect is real, but also makes it virtually certain that it can be resolved by a simple refinement in protocol. Further, scoring probably corrects adequately for this bias.

But the problems with the small a_1's does not contribute to the bias in $\hat{A}$, nor account for all the bias in $\hat{N}$. This much more general inaccuracy must be due to some other factor that produces more grid points falling into the study region, and more hitting a target reach, than are expected. Consider, for example, the 1290 grid dots for the 10 reps, of which a total of 968 hit target reaches. The study region has 7779 sq.mi, of which 5940.8 are in the direct watersheds of target reaches. The nominal grid has 64 sq.mi per point, so that the 10 random placements of the grid were expected to yield 1215.47 hits. The error can be expressed in several alternate ways: 1) the estimated area of the study region is 64*1290/10 - 8256, 2) the estimated grid size is 60.302 sq. mi per point.

The nominal mean estimate of A is 64*968/10 = 6195, and an obvious adjustment to this is 60.302*96.8 = 5837. This amounts to a ratio estimate, with the ratio given by R = 1215.47/1290 = 7779/8256 = 60.302/64 = .9422. Then the estimator is $\hat{T}_R = \hat{T}R$. This is examined more fully in Table 6. The multiplier from statistic to estimate in Table 6 is 6.0302 rather than the 6.4 used in Table 5.

Table 6a Ratio estimates by size class — *all target reaches.*

	DIRECT WATERSHED AREA				NUMBER OF REACHES			
CLASS	n	$\hat{A}_R$	$\sum_P a_i - A$	Δ	$\sum_S 1/a_i$	$\hat{N}_{sR}$	N	Δ
0—.2	3	18.1	2.72	15.4	(6)	36.2	22	14.2
.2—.5	3	18.1	13.76	4.3	(6)	36.2	38	-1.8
.5—1	8	48.2	52.67	-4.5	10.35	62.4	69	-6.6
1—2	65	392.0	305.51	86.5	41.36	249.4	199	50.4
2—3	112	675.4	625.81	49.6	45.72	275.7	253	22.7
3—4	91	548.7	555.46	-6.7	26.27	158.4	160	-1.6
4—5	99	597.0	624.19	-27.2	22.24	134.1	140	-5.9
5—10	327	1971.9	2068.60	-96.7	47.66	287.4	301	-13.6
10<	260	1567.9	1692.09	-124.2	17.95	108.2	114	-5.8
TOTAL	968	5837.2	5940.81	-103.5	247.39	1348.0	1296	52.0

Table 6b Ratio estimates by size class — *headwater reaches, only.*

	DIRECT WATERSHED AREA				NUMBER OF REACHES			
CLASS	n	$\hat{A}_R$	$\sum_P a_i - A$	Δ	$\sum_S 1/a_i$	$\hat{N}_{sR}$	N	Δ
0—.2	0	0.0	.23	-.2	0.00	0.0	2	-2.0
.2—.5	1	6.0	1.03	5.0	(2)	12.1	3	9.1
.5—1	3	18.1	13.68	4.4	3.46	20.9	17	3.9
1—2	48	289.4	211.67	77.8	29.38	177.2	133	44.2
2—3	89	536.7	484.88	51.8	36.11	217.8	195	22.8
3—4	73	440.2	451.07	-10.8	21.20	127.8	130	-2.2
4—5	80	482.4	499.14	-16.7	18.03	108.7	112	-3.3
5—10	256	1543.1	1603.71	-60.0	37.31	225.0	233	-8.0
10<	202	1218.1	1277.22	-59.1	13.38	80.7	84	-3.3
TOTAL	752	4534.0	4542.63	-8.6	161.31	970.1	909	61.2

In Tables 6a and 6b it is seen that estimates over all classes are substantially improved for all four parameters by conditioning on the number of hits in the study region. It is also seen that class by class improvement is not uniform, and some class estimates are greatly worsened. On the whole, however, ratio adjustment seems a good tactic, and worthy of general consideration. Note that the nature of the differences seen in Tables 5a and 5b is such that it would be difficult to achieve an adjustment without worsening some of the classes. Note too, in Table 6, that the average overadjustment of $\hat{A}$ is countered by the average underadjustment of $\hat{N}$, and that adjustment for estimated area of 1st order reaches is very close, with the result that adjustment for higher order reaches is not close.

Some discussion of the nature of the implications of such a ratio adjustment is in order. First, the procedure may be justifiable when the difference in hits and expected hits is no more than random variation. This is the nature of many conditional estimators. Second, the procedure may be indicated if the grid scale and map scale are somehow inconsistent; this could also involve the planimetering process, and any other influence on the measured watershed areas.

It appears from these tables that some additional factor, beyond the inconsistency of scale, is involved in the patterns shown. First order reaches apparently have a slight tendency toward too many in the 1–3 classes, and too few in the greater than 5 classes. Higher order reaches apparently have a tendency for too many in the smallest class, and two few in the greater than 5 class. It would be productive to deliberate over the selection protocol in the attempt to identify causal factors for such tendencies. One is apparent: the occasional very small higher order reach-watershed is an understandable event, and can be eliminated by a benign change in protocol. Other such possibilities may exist.

COMPARISON OF VARIANCE ESTIMATORS

Another analysis made from these data relates to different objectives, and so is incompletely treated here, and will be assessed more completely elsewhere as part of a more general comparison of the Yates-Grundy v-estimator to the Horvitz-Thompson v-estimator, and the overall performance of the Horvitz-Thompson v-estimator, in the manner of the paper by Stehman and Overton, 1987a. These comparisons (Table 7) differ from those ordinarily made in our computer simulations (e.g., Stehman and Overton, 1987a, 1987b) in that the structure of an operational survey is realistically reflected. Specifically, the sample of target reaches is a random subset of the points falling in the study area, and therefore the number of target reaches in the sample is a random variable. We are unable to conduct computer simulations with this structure, and have made those simulations with the number of target reaches a fixed quantity. As a result, an important feature of the Yates-Grundy v-estimator was overlooked until we examined the two v-estimators on the replicated samples, Tables 7a and 7b.

Table 7a. Comparison of Yates-Grundy and Horvitz-Thompson estimates of variance over the 10 replicates — *Number of Reaches.*

	All Reaches		1st Order			>1st Order		
Rep	v_{HT}	v_{YG}	v_{HT}	v_{YG}	v_{YG}*	v_{HT}	v_{YG}	v_{YG}*
1	13,357	6,983	9,651	2,855	5,692	9,073	718	6,479
2	74,376	42,363	12,584	3,136	6,274	72,902	7,723	47,199
3	525,069	356,975	33,691	14,445	20,956	508,347	68,952	352,986
4	638,139	408,952	14,237	4,013	7,561	642,858	96,949	419,995
5	16,303	9,045	16,477	6,037	10,251	4,052	275	3,020
6	30,361	17,545	16,592	5,775	10,196	21,298	2,315	14,878
7	24,942	13,580	10,806	3,276	5,915	21,188	1,576	14,717
8	23,051	11,620	22,479	8,569	12,476	4,281	210	2,853
9	12,041	6,166	13,242	3,990	7,993	2,299	132	1,672
10	12,865	6,308	10,664	3,265	5,963	5,974	467	4,118

*Yates-Grundy variance estimates using y_i=0 for units not in subpopulation, as opposed to restriction of summation to the relevant subset.

Table 7b. Comparison of Yates-Grundy and Horvitz-Thompson estimates of variance over the 10 replicates — *Reach Length.*

	All Reaches		1st Order			>1st Order		
Rep	v_{HT}	v_{YG}	v_{HT}	v_{YG}	v_{YG}*	v_{HT}	v_{YG}	v_{YG}*
1	75,033	23,641	82,614	13,975	43,699	44,155	1,260	31,678
2	116,628	32,421	84,735	9,651	35,422	94,772	3,875	59,878
3	165,324	68,125	117,084	22,576	56,410	116,161	8,068	79,636
4	128,727	46,279	77,395	9,634	34,405	108,439	7,936	68,982
5	69,743	22,522	80,620	13,990	42,193	36,982	1,006	28,187
6	100,903	35,428	106,465	21,383	57,899	54,265	1,710	37,357
7	109,716	36,402	101,546	18,650	48,812	66,736	2,505	46,156
8	118,969	42,316	117,808	28,875	54,919	41,541	1,279	27,777
9	71,256	22,354	88,191	13,339	46,903	28,793	807	21,178
10	78,199	25,023	78,726	14,323	38,978	43,566	1,444	30,140

*Yates-Grundy variance estimates using y_i=0 for units not in subpopulation, as opposed to restriction of summation to the relevant subset.

The critical information in Tables 7a and 7b is that the Yates-Grundy v-estimates are substantially smaller than the Horvitz-Thompson v-estimates. We have identified by analytic methods the reason for this; simply, the Yates-Grundy forms are not generally appropriate for subpopulations. And the greater the population reduction, the greater the underestimate of variance, as seen in the tables.

Specifically, the HT estimates of variance for all reaches account for the subset of grid points that did not lead to a target reach, but the YG estimates do not. Further subsetting to the subpopulations, say, of first order reaches, leads to further evidence of the same effect, and very great underestimation by the YG formula. Changing the way the YG estimator identifies the subpopulation of first order reaches (with v_{YG}*) apparently accounts for this level of subsetting, but there does not seem to be a way to extend this device to the first level effect.

Empirical verification that it is the Horvitz-Thompson forms that are appropriate is provided in Tables 2 and 4. Further verification is provided by computer simulation by randomized *vps* (variable probability systematic) samples of size 75 made from the full list frame of 1296 target reaches. The results in Table 8 are a small subset of the results of those studies that particularly bear on the issue of choice between the Horvitz-Thompson and Yates-Grundy forms, and the subpopulation issue.

Table 8. Selected data from simulated randomized *vps* sampling, from KNOX, with n = 75 and 1500 replications.

Population Parameter	T_y	$\hat{T}_y$	v_{HT}	v_{YG}	$\hat{V}_{sim}$
Total length of all reaches	4192	4191	58426	42627	55339
Total length of 1st order reaches	3184	3170	73620	24101[a]	73327
Number of target reaches	1296	1298	39738	37119	38269
Number of 1st order reaches	909	907	15290	7205[b]	15110
Direct watershed area, all target r.	5941	5941	0	0	0
Direct watershed area, 1st order r.	4543	4529	76244	0[c]	76867

[a,b,c] Yates-Grundy variance estimates using y_i=0 for units not in subpopulation, as opposed to restricted summation are: [a] 69907, [b] 14340 , [c] 76507 .

In Table 8, it is clearly seen that either v-estimator is adequate for estimates of whole population parameters, but that the Yates-Grundy form greatly underestimates the variance of estimates of subpopulation parameters, when subpopulations are identified in the prescribed manner for the NSS, namely by subsetting the data set. The alternate way of identifying subpopulations, by setting

$y_i = 0$ for all reaches not in the subpopulation, is illustrated in the footnote to Table 8. This device is seen to adequately account for subpopulation estimation for these subpopulations, but there are other structures in the NSS that cannot be so represented, as previously indicated.

The results of Table 8 confirm the pattern seen in Table 7, and additionally establish the result that underestimation of subpopulation variance by the Yates-Grundy formula is substantial. This phenomenon is treated in greater depth elsewhere, but our general perception of the result arose in the course of this analysis of the replicated study. This property invalidates the Yates-Grundy v-estimator for many of the objectives of the Surface Water Surveys, since those objectives involve estimation of parameters of subpopulations.

COMPARATIVE PRECISION

It is of interest to compare the results of these studies with theoretical results of other sampling strategies. Specifically, consider simple random samples (SRS) of sizes 97 and 75 from the list frame of 1296 reaches in the full population. Precision of these designs are given in the first two columns of Table 9. The third column presents the comparative precision of the design used in the NSS, as determined from the replicated samples. The fourth column extracts the comparable precision from the simulated samples summarized in Table 8, representing *vps* sampling from the list frame. The last column in Table 9 summarizes estimated precision from the replicates, to be compared with column 3.

Table 9 Standard errors of alternate sampling strategies, and the mean estimated standard error.

	S R S		Reps	*vps*	Mean Estimate
Population Parameter	n=97	n=75	$\bar{n}$=97	n=75	from Replicates
Total number of target reaches	0	0	155	196	160
Number of headwater reaches	58	67	95	123	122
Total area of target watersheds	526	604	241	0	295
Area of headwater watersheds	528	606	259	277	335
Total length of target reaches	264	303	280	235	303
Length of headwater reaches	295	338	275	271	304

The pattern in Table 9 is as expected. A SRS from the list frame would yield greater precision than the variable probability sample from the point/area frame (reps) in estimated numbers of reaches, because the total number of reaches is known from the list, and estimates of subpopulation size benefit from this knowledge. However, the population attribute, direct watershed area, has this same great advantage in the simulated *vps* samples from the list frame, and an edge, but not so great an edge, in the replicates. The parameter, reach length, has almost identical precision in the two designs, but an advantage in the simulated *vps* because of the relation with and constraint on direct watershed.

The pattern is straightforward; only if one is primarily interested in the total number of target reaches is there an advantage in the list frame and SRS. The list frame and *vps* is better for other objectives, and the design used in the NSS (represented in Table 9 by the Reps) is generally comparable, but without investment of effort required by the list frame. Additionally, the point area frame readily provides good spatial distribution of the sample reaches. Note that although the SRS and the simulated samples both use a list frame, the latter requires much greater investment, because the a_i's must be obtained for all units in the population in order to use *vps* sampling from the list frame.

The last contrast of interest in Table 9 is between the mean estimate of SE and the SD among rep estimates. The general pattern of modest conservatism is again apparent. In all cases the behavior of the estimates is adequate.

CONCLUSIONS

From these critical assessments, it would be easy to get the false impression that the sampling and implementation errors under discussion have a serious impact on the NSS results. Quite the contrary is true; relative precision of this survey is very good. As seen in Tables 4 and 9, the errors discussed are reasonably subsumed in the sampling errors, and the estimated standard errors are only slightly conservative.

The purpose of reflecting on these errors, and their patterns, is to better understand the sampling protocol, and to discover additional changes in protocol that will further improve an already quite good sampling design. Several analytic devices to enhance precision are examined.

Scoring reduces the effect of some of the minor errors apparent in application, and additionally enhances accuracy, even if no such errors are present. There is some difficulty in choosing a level of scoring; scoring to 1.0 is appropriate in the population studied here. But some of the subregion populations sampled in the Phase I Survey differ greatly from the study population in distribution of a_1, so a somewhat arbitrary scoring level of .2 was prescribed for the Phase I survey. This is clearly conservative.

The analytic device of scoring involves a very small subset of the sample units (Tables 4 and 5). An alternative device would be a change in the sampling protocol to eliminate these sampling units that have very small a_1's. This is a feasible option, and should be considered for subsequent applications of this sampling strategy. *In fact, future application of this SAMPLE could benefit from reinterpretation of the sample points and the associated a_1's, and such reinterpretation is also a feasible option.*

Ratio estimation offers a potential of reducing a more important bias component in these surveys than does scoring. As many of the analyses are of the nature of ratios, this must be seen as a potentially very useful approach to analysis. Follow up of this direction will proceed as time and commitment permit. Some attention to grid and map scale is also indicated. The need for ratio estimates could be caused by an inconsistency between map scale and grid scale.

An alternate sampling design would begin with the list frame of the full population, as suggested by a contrast made with SRS in the analyses of Table 9. A number of sampling schemes could be used on this frame, most of which would be more precise than the simple random sample. It could be difficult to achieve the spatial representation that the area/point sample gives, but not impossible. The real issue in choice of the area/point frame over the list frame is the effort required to generate the list. There are probably circumstances in which the list frame is advantageous, but the circumstances of the NSS do not seem to be such. In any event, these analyses clearly indicate that the point/area frame used in the NSS is generally comparable in precision to others that might have been used, and has clear operational advantages.

Several of the reviewers of the stream sampling plan were concerned about proposed use of the Horvitz-Thompson variance estimator, which has been shown to behave badly in some circumstances. While it is true that in part the good performance of this v-estimator derives from use of the novel approximation for the pairwise inclusion probabilities, these studies have shown that the HT v-estimator does not exhibit these poor attributes in the circumstances of the stream survey, even when the Hartley-Rao approximation to the π_{ij}'s is used (Stehman and Overton, 1987a). Not only is the HT v-estimator adequate, but it alone is adequate for some of the objectives of the NSS. Additionally, it is computationally more convenient than the Yates-Grundy formula in the context of the NSS.

ACKNOWLEDGEMENTS

The data for this study were collected by Andrew Kinney, Anastasia Allen, Douglas Brown, Darrell Downs, Jeffrey Irish, and Suzanne Pierson. Rob McLeod and Greg Weier provided innovative data management programming. Analyses were made by the authors.

REFERENCES

Allen, A. B., and Kinney, A. J. (in progress) National Surface Water Survey: A Regional Approach for Stream Selections. EPA Report.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 47, 663-685.

Messer, J.J., C.W. Ariss, J.R. Baker, S.K. Drousé, K.N. Eshleman, P.N. Kaufmann, R.A. Linthurst, J.M. Omernik, W.S. Overton, M.J. Sale, R.D. Shonbrod, S.M. Stanbaugh, and J.R. Tutshall, Jr. (1986). *National Surface Water Survey: National Stream Survey, Phase I — Pilot Survey.* EPA—600/4—86—026, U.S. Environmental Protection Agency, Washington, D.C.

Overton, W. S. (1985). *A Sampling Plan for Streams in the National Stream Survey.* Technical Report 114, Department of Statistics, Oregon State University, Corvallis, Oregon, 97331.

Overton, W. S. (1987) *A Sampling and Analysis Plan for Streams in the National Surface Water Survey.* July 1987. Technical Report 117, Department of Statistics, Oregon State University, Corvallis, Oregon, 97331.

Stehman, S. V., and Overton, W. S. (1987a). Estimating the variance of the Horvitz-Thompson estimator for variable probability, systematic samples. *Proceedings of the Survey Research Section*, 1987 *ASA Annual Meetings*, San Francisco.

Stehman, S. V., and Overton, W. S. (1987b). *An Empirical Investigation of the Variance Estimation Methodology Prescribed for the National Stream Survey: Simulated Sampling from Stream Data Sets*. Technical Report 118, Department of Statistics, Oregon State University, Corvallis, Oregon, 97331.